CSE 578: Climate Change Impact Visualization Using Global Temperature Data

Sarthak Bharatkumar Patel
Arizona State University
ASU ID: 1229983523
spate271@asu.edu

Jay Amrutbhai Patel
Arizona State University
ASU ID: 1228156503
jpate132@asu.edu

***Abstract*—In our modern society, data plays a pivotal role in enriching our lives, informing decisions, and solving challenges. Through CSE 578: Data Visualization, the document introduces the project titled "Visualization of Climate Change Impacts Using Global Temperature Data". The goal is to develop visual representations that demonstrate the effects of climate change using historical global temperature data. By analyzing temperature records spanning a century, the project seeks to depict trends, anomalies, and instances of extreme weather events through time-series analysis and forecasting evaluation. The primary focus is on emphasizing the critical nature and consequences of climate change to get ready for the upcoming temperatures.**

***Keywords*—Climate Change, Temperature Trends, Global Warming, Data and Geographic Visualization**

I. Introduction

This Global warming is one of the major issues that threaten our world today, having significantly impacted and continued to impact the environment, health, and economic systems in the globe. New patterns of weather and climate including more frequent and severe storms, higher sea levels, and changes to the character and function of natural systems are directly related to the increase in global temperature brought about by human activities such as the use of fossil fuels and deforestation [1]. Knowledge and illustration of the trends in global temperature over the years are crucial in passing information and informing policy-making to deal with the effects of climate change.

The purpose of this project is to present the historical account of the global temperature trends over the past century. The primary purpose is to develop visualizations that highlight the effects of global warming using a large dataset of global temperature records. These visualizations will depict temperature trends, anomalies, and severe weather occurrences indicating the severity of climate change and its necessity. To achieve this goal, the project utilizes data sourced from Berkeley Earth which possesses detailed information about the temperature records from various city sources across the world. The methodology involves several pivotal stages: data collection, data preprocessing, exploratory data analysis, data visualization and forecasting. Using statistical tools, techniques and analyzing time series, this project seeks to determine any anomalies in temperature fluctuations in separate time spans and then visually represent such fluctuations over several decades.

II. Description of Solution

This project was mainly based on providing an insight into climate change hazards using global temperature data available from history. The purpose of our study was to create visualizations to reveal temperature trends, trends in variation as well as instances of extreme weather conditions in the past hundred years. Hence, our objective was to demonstrate complex data with strong analytic tools and quantitative analysis in an understandable manner so that the audience varying from research workers to ordinary people, could easily understand it.

A. Data Collection

A crucial part of our project was choosing the Berkeley Earth Surface Temperature dataset [2], an extensive collection maintained by Lawrence Berkeley National Laboratory as it was the only dataset which was containing the accurate information even from the airport records as before hundred years, there were no precise machines in every city. This dataset incorporates 1. 6 billion temperature records gathered from 16 previous datasets originating from measuring processes which begun in the mid-1750s. The cleaning and transformation of data were very rigorous because of changes in measurement metrics due to its historical evolution say from use of mercury thermometer to electronic ones. From this dataset, files used include; “GlobalTemperatures.csv”, “GlobalLandTemperaturesByCountry.csv”, “GlobalLandTemperaturesByCity.csv”. The “GlobalTemperatures.csv” file contain detailed records of Global Average Temperature for both Land and Ocean surface with the maximum and minimum temperatures and one standard error uncertainty for every record. “GlobalLandTemperaturesByCountry.csv” provides average land temperatures represented through countries to assess the temperature changes at a regional level. On the other hand, the dataset “GlobalLandTemperaturesByCity.csv” provides fine-grained temperature values at the city level which helps us in studying the Urban heat island effect or fluctuations in temperature in specific cities. The structure of this dataset was well documented, and the coverage was very detailed for our needs, enabling an accurate portrayal of the variability of temperature throughout history and the consequences of climate change throughout various parts of the world and time frames.

B. Data Cleaning & Pre-Processing

Data cleaning and processing were some of the most important preliminary processes followed in the analysis. Handling null values was done before feature scaling by deleting the rows that contained the missing temperatures. We improved on temporal analysis by deriving the year, month, and month name units from the date fields. In addition, the given coordinates of latitude and longitude were improved following directional indicators such as N, S, E and W direction indicators for better matching with the data tools.

In our analysis, we selected data for the period ranging from 1912–2012, which is considered as the most central time for modern climatology investigations as the last century. In this regard, we made efforts to clean all the collected data in order to avoid the inclusion of irrelevant and unrelated information. To follow this up and enrich our understanding, average temperature of the ocean was found using the formula; average temperature of land + average temperature of ocean –average temperature of land (as obtained from the global temperature datasets). These preprocessing measures ensured that most of our data was free from errors and in a form that could be optimally analyzed and visualized.

C. Data Visualizations

In our project, the visualization of data was critical in our understanding and analysis of the effects of climate change. We used several visualizations in order to present temperature trends, deviations, and illustrations of extreme weather conditions. Using time-series study, we looked at differences extending over many years, as represented by geographic maps that highlighted spatial characteristics or effects.

Initially, we computed annual global average temperatures for both land and sea using Matplotlib to visualize trends from 1912 to 2012. The line plot revealed distinct patterns: land temperatures showed a significant upward trend, suggesting a clear warning signal. In contrast, ocean temperatures exhibited more complexity due to intricate heat exchange processes. This analysis provided insights into global warming by illustrating temperature changes across land and ocean over the century.

We created an interactive Dash tool for detailed analysis of global temperature fluctuations, focusing on India. It includes mean and standard deviation lines to highlight deviations from typical temperature norms, especially in key months. Users can dynamically explore temperature variations, gaining insights into climate changes over time and identifying periods of extreme anomalies. This tool is essential for understanding climate impacts on specific regions like India and informing targeted adaptation strategies in response to climate change.

For example, due to variety in regions, we used Dash and Plotly to illustrate the geography aspect. This translated into presenting the average temperature of each country by year from the year 1912 to the year 2012 because users can easily observe the change in climate in various parts of the world. It helped to make a clear comparison of the changes in temperature using choropleth map specifically for the countries that showed the highest degree of variation. This geographical perspective was crucial for illustrating the uneven impacts of climate change, showcasing regions where warming has been more pronounced.

Furthermore, we created a vertical box plot for distribution of the average land temperature by months. This plot made it easier to understand the cyclic changes that occurred over the year and throughout the century, as it highlighted how individual months affected the general temperatures. In this case, when comparing the distribution across months, they would show patterns such as; high summer temperatures or change in seasonal cycles.

Another important visualization that showed the temperature dynamics of two different time intervals (1912-1916 and 2008-2012) in several countries. This was presented in the form of a bar plot that showed the rise in temperatures in different years while especially laying attention on the average temperatures of the early 20th century and the early 21st century. This comparison pointed to the fact of global warming rate and the countries that have been experiencing higher levels of temperature.

Altogether, these visualizations provided a bird's eye view of global temperature patterns, changes and regional impacts globally. They underscored both the topical nature of the problem and its importance by coloring the climate change data in a simple yet effective way. In our teaching, we used a set of active learning tools and scales, which helped enrich and illustrate the necessary concepts and ideas about climate change and its far-reaching consequences.

D. Time Series Analysis

Our time series analysis was geared towards the examination of temperatures within Tempe, United States whereby data modeling was conducted using the ARIMA model. First, we preprocessed the dataset in a way that only contains temperature data belonging to Tempe. After the cleaning and preparation of the data, the data is indexed by time or date.

To estimate the best set of parameters for the ARIMA model, a grid search was conducted with different values of p, d, q being tested. To divide the dataset the system used training and testing sets and 60 and 40 percent of the dataset were used for training and testing respectively. The models were initially built for the training set using a technique known as differencing and auto-regressive integrated moving average (ARIMA) was used for all the variables to identify the best model using the Akaike Information Criterion (AIC). When the models with the lowest AIC score were determined, the best-fit model for the analysis was chosen.

As a final step, the chosen ARIMA model was then further calibrated on 90% of the data to enhance the model's ability to predict the rain pattern. Employing this model, we generated the forecasts of the temperatures for the testing set. Evaluating its performance was done by comparing the forecasted temperatures to the actual ones, and a plot showing the training data, testing data as well as the forecasted data was drawn. For the purpose of evaluating the accuracy of the prediction, the Mean Absolute Error (MAE) was calculated. The analysis of this time series provided useful information about the analyzed temperature conditions in Tempe, which indicates the general applicability of ARIMA models for climate data analysis.

III. RESULTS

A. Global Temperature Trends From Line Chart

The analysis of the data on the global average temperature revealed a significant change in the land temperature. On the same note, the history of ocean temperatures was also more complex though the chart illustrated an overall positive trend within the same period. From this apparent warming trend, it is easy to see that climate change is indeed affecting the climate conditions both on land and in the seas positively.

B. Country-Specific Analysis with Limits

The engaging visual display primarily concerned with India showcased temperature patterns and highlighted time frames associated with noteworthy climate events. This visualization used mean and standard deviation lines to identify specific years that showed deviation from normality in a given state. It is critical to consider the findings such as the area's coverage for a better understanding of the impacts of climate change.

C. Geographic Variations

The choropleth map visualization represented how the intensity of positive temperature alteration differs from one region to the other at different time periods. As a consequence, communities living in the Northern countries experienced a more drastic rise than those living in the equatorial areas, making it important for the government and international organizations to analyze the findings when designing climate change policies and programs for those areas.

D. Monthly Temperature Distributions using Box Plots

The vertically oriented box and whisker plot for the month-wise average temperature for this piece of land, that was provided, displayed seasonal trends and the general trends. In fact, the temperatures for summer months were comparatively higher, with temperatures increasing steadily over the course of the CENTURY. The above data is very useful in managing and preparing the calendar for seasonal changes that are likely to affect agricultural activities that result from climate change.

E. Historical Comparison of Temperature Changes

The bar plot based on the temperature variations in top countries and the two periods of 1912- 1916 and 2008 –2012 established significant increase in temperature particularly in the northern areas of Greenland and Russia. It also highlights the trend of warming and corroborates the rate of warming which is being experienced in certain regions of the world.

F. Time Series Analysis (ARIMA)

The parameters applied on Tempe, USA collected data enabled future temperatures to be well predicted judging from the low statistical MAE value of ARIMA. This thus illustrates in relatively proof of the regularity determinations of the actual model in terms of temperature fluctuation. Sophisticated time series may prevent long-term strategy to maintain advisory for emerged climate conditions.

G. Extreme Weather Events

Comparing more years of average temperature in India helped provide historic perspective in relation to the frequency of such temperature extremes. Analyzing trends in various types of climatic subjugations is very vital in recognition of prevention measures and the formulation of adequate measures to ensure helpless communities are protected.

H. Utility for Stakeholders

As the paper presents an exhaustive list of the available visualizations and analyses it provides important information for researchers, policymakers, and the general populace. Such resources could be useful for decision making and making changes, raising public awareness about climate change, and could make educational initiatives more effective when it comes to tackling climate change.

Altogether, these findings underscore the importance of using data analysis techniques for modelling and mitigation of climate change. Thus, providing an accurate and detailed picture of past and expected temperature fluctuations, this project ensures the relevance of projects aimed at creating a viable and sustainable world.

IV. MY CONTRIBUTIONS

A. Data cleaning and pre-processing:

With detailed discussions of comparative datasets like NASA GISTEMP and NOAA's MLOST, the Berkeley Earth Surface Temperature data was selected as the most optimal solution due to the vast and exhaustive data points it provides. Temperature recorder data from multiple sources for different time registers this data set from the early days of mercury thermometer to the advanced electronic probe. Both extensive annotation on the data and strict procedures for the transformations of the data as was enacted by Berkeley Earth allowed for rigorous data sanitization. We also beautifully managed missing or null values by removing it, preprocessed date format to easily access while performing the operations, calculated similar metrics such as ocean temperatures to be parted from the column and lastly selecting the relevant time frame from 1912 to 2012. I consider this to be crucial in developing a set of good quality data ready for further investigation and presentation so that global temperature trends and their effect on climate change information could be properly investigated and understood.

B. Global Temperature Trends From Line Chart

The graph showing temperature change over the period 1912 to 2012 shows different trends for land and ocean surface temperatures. In the centennial timespan, it is evident that the red curve associated with the average temperature on the territory is on the rise, which points towards a definite warming process. Although, in terms of variations in the temperature of the ocean, the blue line which is representative of the average ocean temperatures is more undulating suggesting that there were regular fluctuations in the temperature of the ocean during the same period. Thus, this visualization continues to support the finding that climate changing has not ceased to manifest, as land temperatures continue to rise steadily, while ocean temperatures, despite showing values that are measurably rising, vary more due to factors like ocean currents and heat uptake capability. These patterns are important to consider in order to predict the effects of climate change on terrestrial and aquatic environments, as well as to design necessary measures of prevention and control.

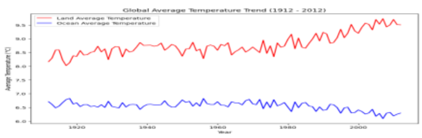

Fig 1: Global Temperature Trends over the century

C. Monthly Temperature Distributions using Box Plots

The analysis based on the Berkeley Earth Surface Temperature shows the distribution of the land average temperatures per different months and it follows to highlight the seasonal changes according to the vertical box plot. There is also a notion that the weather rises in July as the plot has a minimum value in January, which defines the difference between seasons. From this understanding of temperature dynamics, one can learn how the temperatures differ from one month to another with July exhibiting highest temperatures and January showcasing low temperatures. This information remains beneficial in analyzing the seasonal changes in climate patterns and can help to design agricultural coherent schedules and power consumption or coping with the consequences of climatic fluctuations. The plot underlines the ability of visual tools, more specifically, box plots to analyze temporal trends in the data, which is inclusive of climate data, and, consequently, improves our understanding of changes in variability.

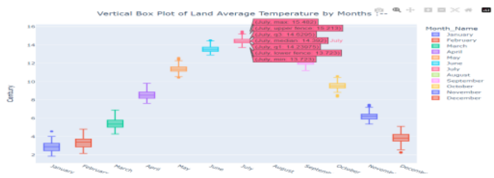

Fig 2: Vertical box plots of temperature over months

D. *Time Series Analysis*

Specifically, when it comes to performing time-series analysis on the temperature data for Tempe, USA, the ARIMA (Autoregressive Integrated Moving Average) modeling approach was selected for use, given that this is a technique that is ideal to integrating temporal factors and trends in data. The model selection when presuming ARIMA involved comparing different values of p, d, q in a grid search procedure destined for the selection of a model with the lowest value of AIC, a measure of model fitness. By having a look at the accuracy of the model on the test data, we can confirm that it was able to follow the different patterns in the temperature data to generate reasonable and accurate future temperature forecasts with the chosen ARIMA model. The plot shows how the temperatures as seen from the actual data of train and test data mimic the forecasted temperatures which show how the model follows the observed trend with increasing time. It is seen that the Mean Absolute Error (MAE) value is approximately equal to 1.4°C Forecasts of the temperatures under consideration are also almost equal to the observed temperatures, which is a clear indication that the model developed under the ARIMA approach is well suited for temperature forecasting for Tempe. As shown in this analysis, ARIMA has general application for the analysis of climate data and can be particularly beneficial in guiding strategic planning for organizations and communities seeking to adapt or mitigate the effects of climate change.

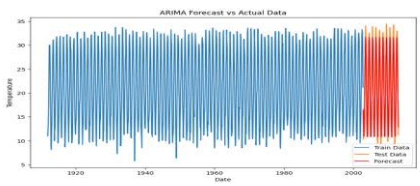

Fig 3: Forecasting v/s Testing Data

V. LESSONS LEARNED

The following are some of the most relevant observations to be made following the completion of this project: Breaking down climate data is challenging and is one of the most important processes:

- First of all, the most critical step is the choice of a wide and reliable sample. The Berkeley Earth Surface Temperature dataset was particularly successful because of its broad inclusion and thorough elaboration. Combining temperature records from different data sources and different years helped to create a stable framework for the analysis done with this dataset and allowed for the use of accurate and relevant data.
- Second, data cleaning and preprocessing stage were more complicated than expected because of the existence of missing or incomplete data. Historical climate data may also carry a lot of missing values or even contain values that are not applicable making the data cleaning phase more important. This step was important in clearly showing that there is need for precision when one is dealing with larger set of data.
- Third, the use of statistical measures like the mean and standard deviations of temperature records was also useful in evaluating possibly abnormal and persistent climatic changes. These statistical techniques provided a systematic approach for examining the areas that deviated from standard value which are important in analyzing the effects of climate change.
- Fourthly, extending the arm with interactive visualization tools like Dash to enhance users' experience improved it greatly. For this purpose, these tools helped in engaging and interacting with more dynamic and richer climate data, converting it into more effective and easily comprehensible information. This interactive capability is useful to directly engage with the stakeholders and help them to make informed decisions on problems faced by organizations.
- Finally, the analysis also concluded on the issues of understanding the trends as far as the temperatures are concerned across the regions. Sources for global warming are varied across the world and so are their outcomes, this is perhaps evidenced from our geographic analysis which was the foundation of this comparison. This understanding is incredibly vital in order to have specialized climate adaptation planning to direct the resources to places that are severely impacted by climate change.

Altogether, these lessons emphasize the importance of choosing the samples wisely, pre-processing the data carefully, analyzing the results meticulously, using visualizations to make the findings more engaging, and working with the regional data to contribute to the development of climate change research.

Team Members: Sarthak Bharatkumar Patel, Jay Amrutbhai Patel

REFERENCES

[1] "Data Overview – Berkeley Earth," Berkeley Earth. https://berkeleyearth.org/data/

[2] National Oceanic and Atmospheric Administration, "Climate Change Impacts," National Oceanic and Atmospheric Administration, Aug. 13, 2021. https://www.noaa.gov/education/resource-collections/climate/climate-change-impacts

[3] How to Create an ARIMA Model for Time Series Forecasting in Python – Machine Learning Mastery https://machinelearningmastery.com/arima-for-time-series-forecasting-with-python/